大数据时代知识管理过程、技术工具、模型与对策*

■ 叶英平 陈海涛 陈皓
吉林大学管理学院 长春 130022

摘要:[**目的/意义**]大数据时代给传统知识管理带来了变革,为适应大数据时代的到来,需要重新认识和定位知识管理过程,应用大数据技术工具构建新的知识管理模型,以应对知识管理的更新与挑战。[**方法/过程**]通过梳理数据到知识的形成过程,结合大数据的 4V 特征,了解知识管理所需的大数据技术工具,并将大数据时代的知识管理划分为知识生产、知识积累、知识交流、知识应用 4 个阶段,结合大数据技术工具的使用,构建新的知识管理模型,提出应对碎片整合、应用价值、硬件支撑、隐私伦理等问题的对策。[**结果/结论**]大数据的发展推动了知识管理过程从传统模型向技术模型的转型,大数据时代的知识管理模型与大数据技术的使用紧密相关,强调从海量碎片化数据中提炼知识价值,并更有效地辅助组织决策,为此,需要做好硬件设施支撑和信息安全保障,将大数据技术与小样本分析相结合,推动知识管理走向新的层次。

关键词:大数据时代 技术工具 知识管理模型
分类号:G203
DOI:10.13266/j.issn.0252-3116.2019.05.001

1 引言

近年来互联网行业发展风起云涌,而移动互联网、电子商务、智慧物联以及社交媒体的快速发展更促使我们进入了大数据时代。关于大数据的定义往往根据其特征来描述。2001 年,Gartner 公司分析师 Doug Laney 最先提出了大数据"3V"模型的概念,包括数量(volume)、速度(velocity)和种类(variety),并将大数据概括为海量、高增长率和多样化的信息资产。此后,不同学者在"数据量大"这一特征的基础上,从时间、空间和多样性 3 个方面阐述了大数据的内涵与外延,增加了真实性(veracity)维度,将数据质量纳入 4V 的范畴,刘智慧[1]在浙江大学学报上对大数据技术的相关研究进行了综述,并对大数据的 4V 特征进行了全面系统的阐述。大数据的 4V 特征一经提出,被学术界普遍接受,并在社会上获得了广泛认可。IBM 公司结合大数据的市场应用,在原有框架内为大数据增加了价值(value)属性,继而从价值流转的角度,发展出连通性(valence)与可变性(variability),将大数据的特征由 4V 拓展 5V、7V 甚至 11V(动态性 vitality、可视化 visualization、合法化 validity、有效性 validity)。

伴随着互联网技术及计算机科学的发展,几乎所有领域都受到大数据时代的影响,面临转型升级的迫切需要。尤其是对于以数据和信息为研究对象的知识管理方向来说,如何高效便捷地发现新知识、提供新产品、创造新价值,从而提升企业主体和社会组织的竞争能力,引起了国内外学者的广泛关注和高度重视。

为使海量碎片数据的信息价值能够得到充分使用,知识管理的作用日渐凸显。已有学者在传统知识管理的基础上,借由大量数据扩充知识来源,并以此拓展了知识管理的外延,如 P. Cooper[2]、赵蓉英等[3],但未能深入挖掘大数据时代的特征内涵,也有学者对大数据时代的技术工具进行梳理和介绍,如孟小峰等[4]、许立波等[5],但没有结合大数据技术工具的使用,提出变革性、创新性的知识管理模型。此外,现有研究涉及的大数据时代知识管理的主体,多为图书馆、档案馆等传统知识服务机构,如刘洁璇[6]、刘捷[7]、曾润喜等[8],

* 本文系中国科学技术协会高端科技创新智库青年项目"产学合作中网络权力、网络惯例对企业创新能力的影响机制研究"(项目编号:DXB-ZKQN-2017-030)和中国博士后科学基金面上资助项目"创新联盟中网络惯例的演进及对联盟绩效的影响机制研究"(项目编号:2018M641789)研究成果之一。

作者简介:叶英平(ORCID:0000-0003-0577-8025),讲师,博士后,E-mail:351614247@qq.com;陈海涛(ORCID:0000-0003-4198-7318),教授,博士生导师;陈皓(ORCID:0000-0002-7092-5506),硕士研究生。

收稿日期:2018-07-08 修回日期:2018-11-05 本文起止页码:5-13 本文责任编辑:易飞

对于大数据发展所带来的决策支持功能和知识增值过程则研究不足。为此，本文在梳理“数据－信息－知识”形成过程的基础上，分析了由传统知识管理向大数据知识管理的转变，结合大数据的 4V 特征介绍大数据时代的知识管理工具及知识管理方式，构建基于大数据背景的新的知识管理模型，并从数据碎片整合、信息安全保障、决策支持辅助、硬件设施支撑 4 个方面提出相应的对策措施，从而在大数据环境下更好地实现知识管理的目的，即通过知识管理提高组织的竞争能力与创新能力，进而推动社会化生产效率的提升。

2 大数据背景下知识管理过程分析

对于数据与知识的关系，P. Cooper[2] 认为在特定环境中数据的描述含义构成了信息，而对这些信息进行结构化和组织化则形成了知识，可以说，数据是一种基础信息，而信息经过处理提炼可构成知识。在大数据背景下，大量的碎片化的原始数据是分散而孤立的，经过处理与解释，使数据之间建立了关联、形成了信息，而信息经过加工与分析转化为知识，并应用于生产生活。数据与信息的囤积是没有意义的，只有通过数据挖掘与知识管理，使得数据信息服务于经济社会发展，才能发挥出大数据的真正价值。赵蓉英等[3] 认为，所谓知识即是对数据和信息进行提取、整合与固化的结果，并指出在大数据的 4V 特征下，需要对数量巨大、来源广泛、格式多样的数据集合进行采集、存储与挖掘，应注重提炼数据价值，实现数据－信息－知识－智慧的转变。

结合上述学者的研究梳理可知，在大数据背景下，不同渠道产生的大量碎片数据经过提取、清洗、转换，进入数据仓库形成特定数据集合，再将多个数据集合通过线上网络相联，构建多维数据立方体，进行动态分析并整合集成为信息资源，通过描述性分析与预测性分析，挖掘信息模式及潜藏规律，归纳概括为隐性知识，最后通过可视化将这些知识显性化，以直观的可视化画像呈现出来，并由知识间的关联性形成知识图谱。由此推导知识的形成过程如图 1 所示，从数据向知识的转变，历经了数据仓库、联机分析、信息挖掘、知识呈现 4 个过程，伴随着数据流、信息流、知识流、价值流的形成。所有数据、信息、知识的价值，最终通过显性化表达的知识图谱得以实现。此外，考虑到大数据时代海量数据的交叉融合，使得知识的形成过程不是数据到知识的一对一关系，而是多对多的形成过程，因而在知识图谱的构建过程中，也存在相应的数据网、信息网、知识网和价值网。

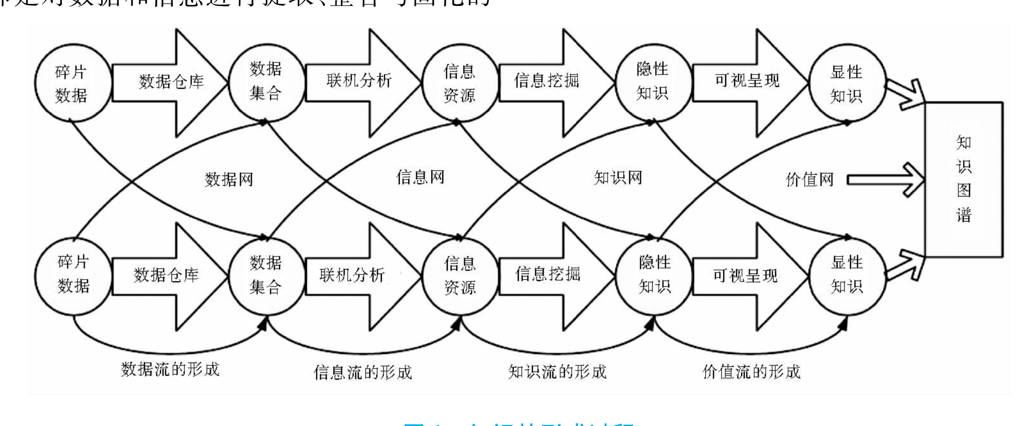


图 1 知识的形成过程

知识管理是企业（或其他机构）对内部知识资产进行的管理，其核心任务是将显性知识与隐性知识相互整合，通过知识获取、吸收、转移等过程，实现企业（或其他机构）的知识创新与价值创造。为此，日本学者野中郁次郎等[9] 提出了 SECI 模型，认为隐性知识经由社会化（socialization）、外在化（externalization）、组合化（combination）、内隐化（internalization）4 个过程，完成了知识管理的螺旋式上升。知识管理的关键在于知识转移，众多学者从增加知识转出方转移意愿、增强知识接收方吸收能力、适当缩短知识距离等方面进行研究，力争扩大知识搜寻范围、优化知识编码方式，从而提高知识转移效率、提升知识整合效果。

随着大数据时代的到来，学者们将对知识管理的关注焦点，由隐性知识向显性知识的转化，迁移到如何从海量多源异构的数据中挖掘知识价值，为此，国内外学者围绕大数据在知识管理中的理论发展及技术应用，展开了一系列研究。在理论研究方面，A. Simonet 等[10] 基于大数据增量的敏捷性，引入了活性知识的概

念,认为大数据的灵活性能够有效提高知识管理的表现能力;秦殿启[11]将大数据理念运用于知识整合理论,基于个人组织的层面,论述了知识管理的模式、特点和具体步骤;在应用研究方面,C. S. Liew等[12]阐述了如何利用大数据环境下的移动智能终端平台,挖掘数据中的有效信息,并进行知识管理;黄天恩等[13]设计了基于大数据的仿真模型,运用知识管理的方法进行电网超前预警,解决了能源互联网环境下的安全一体化问题;何军[14]基于大数据背景下决策环境的变化,系统阐述了大数据对企业知识管理和战略决策的影响。

学者们普遍认为,大数据环境下的知识管理需要快速处理更大体量的数据集,并以更优化的知识管理方法挖掘海量低密数据的知识价值。叶英平等[15]研究了基于网络嵌入的知识管理模型。与传统的知识管理研究相比,大数据在知识获取、知识存储、知识整合、知识使用等知识管理过程均有所不同:①知识获取由传统知识管理面向组织外部的定向搜寻和主动获取,转变为海量碎片化数据的涌现,扩大了知识获取的范围和体量,改变了知识获取的方向;②知识存储由私有知识库存储的方式向共享分布式存储的方式转变,这种云端存储的方式,扩展了知识存储容量,突破了知识存储所有权力的限制;③知识整合由传统知识管理中结构化、可编码的数据整合,向结构化、半结构化、非结构化并存的多源异构数据的逻辑整合转变,增大了知识整合的难度,优化了知识整合的方式;④知识使用的关键前提条件由隐性知识的显性化,转变为数据清洗与数据脱敏,大数据时代的信息共享使得对于数据处理结果的应用,由服务于某一主体的知识增值,到被多方主体使用。根据传统知识管理向大数据背景下知识管理的转变,可以概括出大数据背景下的知识管理过程如图2所示,即随着时间推移,逐渐深化知识的加工程度,通过海量碎片数据的收集完成知识获取,以分布式方式存储知识,通过处理多源异构数据完成知识整合,最终将数据分析结果应用于商业价值的实现。

3 大数据背景下知识管理技术工具

不同于传统知识管理过程,大数据背景下的知识管理需要借助更多高科技工具与软件,各类大数据实用技术的使用推进了知识管理过程的变革,进而推升了知识管理水平和组织运行效率。孟小峰[4]在研究大数据时代计算机技术发展的同时,介绍了诸多大数据

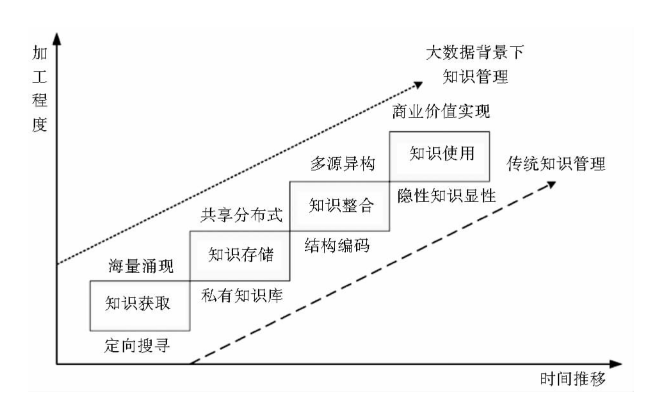


图2 知识管理过程

处理工具,包括用户生成内容、感知式系统、流处理与分处理模式、人机交互技术、Caffeine索引、Pregel图计算模型等。根据大数据数量庞大、动态增长、多源异构、价值低密的4V特征,大量迅速增加的数据集合,需要借助数据清洗工具实现知识获取;海量半结构化和非结构化的数据,需要借助数据仓库工具实现知识存储;纷繁复杂而又质量参差的数据,需要借助云计算工具实现知识整合;动态增加的原生数据,需要借助可视化工具进行数据挖掘,从而提炼知识价值并被使用。据此,知识获取、知识存储、知识整合和知识使用4个知识管理过程中所涉及的数据清洗、数据存储、数据分析和数据挖掘工具如图3所示:

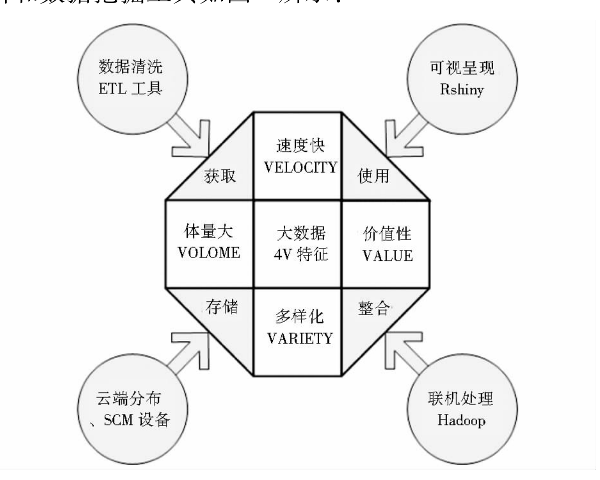


图3 知识管理的技术工具

3.1 数据清洗技术及工具

大数据的庞大体量并未增加数据价值,反而意味着数据噪音的增多和信息垃圾的泛滥,因此,在获取海量数据之后,首先要进行数据清洗这一预处理工作,将非清洁数据降噪转化为高质量的干净数据,从而更好地进行后续数据分析与知识管理。这种数据的非清洁性一方面体现为数据本身的不准确、不完整,另一方面体现为数据之间的冗余和不一致。彭宇等[16]认为数

据清洗是指对所需数据源进行抽取集成,并从中提取关系实体,关联聚合后统一定义结构的过程,旨在提高数据质量,为数据分析打好基础。程学旗等[17]也指出,数据清洗是数据分析的前提,包括基于物化或ETL、基于联邦数据库或中间件、基于数据流、基于搜索引擎四种方法。大数据时代细微有用的信息混杂在庞大的数据量中,因而数据清洗过程需要在质与量之间进行权衡,粒度过细可能会过滤掉有用信息,而粒度过粗又无法达到清洗效果。借助 ETL(extraction,transformation,load)工具,可以通过提取 - 转换 - 负载的过程,将来自不同运行平台、编制语言、物理位置的数据按照统一格式提取出来,再进行清洗、转换、集成,加载进入数据仓库。

3.2 云端分布技术及工具

大数据的出现对于数据存储的颠覆不是简单的技术演进和空间扩容问题,小到数据来源、大到数据思维,都为数据存储带来革命性变化。由于存储空间的需求量巨大,需要采用云端分布的数据仓库进行数据存储。大数据的动态增长对于数据仓库响应时间的要求提出了考验,区别于传统知识管理中静态关系数据库的概念,数据仓库还涉及了决策支持系统中大数据的动态存取问题,需要在数据仓库中对大数据进行数据处理。金澈清等[18]将传统数据库与新兴数据存储技术进行了对比,认为在存储基础上,数据仓库还具备数据处理功能,具体划分为流处理模式与批处理模型两种方式。其中,流处理更多依赖于内存中的概要数据结构即内存容量,需要借助 PCM(相变存储器)等储存级内存(SCM)设备,如推特公司的 Storm、雅虎公司的 S4 以及领英公司的 Kafka 等;而批处理模式的代表则是谷歌公司的 Map Reduce 编程模型,通过将数据源进行分块,交给不同的 Map 任务区,把解析出的链/值(key/value)对集合执行 Map 函数,将得到的中间结果写入本地硬盘,再由 Reduce 任务根据 key 值进行排序,并输出最终结果。

3.3 联机处理技术及工具

邱东[19]从统计学的角度,分析了大数据环境对数据处理提出的挑战。而联机处理为大数据提供了弹性扩展、成本较低、模式丰富、资源庞大的计算能力,可以同时处理结构化、半结构化、非结构化等形式的异构系统,通过网络将计算能力组织起来,实现数据处理的规模化和专业化。随着 4G 网络的兴起和移动终端的使用,还衍生出移动云端联机处理的概念,将网络中的存储数据借由应用软件集合起来,协同计算,从而为知识发现打下基础。联机处理工具如谷歌公司开发的 Hadoop,是一个能够对大量数据进行云计算的软件框架,作为目前最为流行的云计算开源平台,支撑了谷歌公司内部的大数据处理与知识管理应用;又如微软公司使用 Dryad 模型构建的支持有向无环图(DAU)类型数据流的并行程序,通过联机处理过程可实现提炼数据价值的目的。

3.4 可视呈现技术及工具

可视化技术是数据挖掘和知识发现的必要手段,是促进知识传播和转化的重要工具,也是知识管理系统的重要技术基础。通过可视化技术,丰富了知识的表现形式,降低了知识的理解吸收难度,激发了人们接受和学习知识的意愿,促进了知识在组织内外的传播共享。任磊等[20]认为可视分析以总结性表达的方式,更好地实现了对知识库的调用与掌控,有利于知识的更新迭代,并在研究中介绍了常见的可视化技术,包括标签云、历史流、空间信息流等,以及 Rshiny、Flot、Raphael、Google Charts、SAS Visual Analytics 等可视化软件。在实际应用中,可以根据数据挖掘与知识发现的需要,选择合适的可视化技术和可视化程序,并通过可视化,实现数据价值的使用。官思发等[21]指出,联机处理过程和可视化呈现对于数据本身的真实性要求较高,只有相对准确、可以信赖的高质量数据,其计算处理与可视分析的结果才具有意义与价值,如果数据本身存在错误或问题,那数据分析和数据挖掘的结果也不会是正确的。同时,可视化技术还需要借助元数据的管理,元数据是关于数据的数据,体现了数据之间的关联关系以及数据本身的固有属性,元数据的表征有利于对数据进行可视化呈现展示。

4 大数据时代知识管理模型构建

在大数据时代,数据从传统知识管理的处理对象,转变为知识管理的基础资源,这些实时动态数据、非结构化数据乃至机器生成数据,造成了数据庞大体量的规模性和多源异构的复杂性,并为传统知识管理的知识存储与知识整合带来了挑战。大数据环境下的知识管理,体现了知识量激增、知识处理速度快、知识更加多样化、知识价值密度降低而知识价值总量增加、知识处理更加信息化和智能化等特点。数据量的增长并非将知识管理系统简单扩容就能解决,量变引起质变,传统的知识管理理念和技术已经无法应对海量的大数据时代,知识管理系统的更新速度与处理速度已经超过了传统知识管理的提速方式和逻辑负荷,需要以新兴

的大数据技术重塑知识管理流程，以创新的知识管理方式规整碎片化大数据。

借鉴传统知识管理中知识获取、知识存储、知识整合、知识使用的过程，大数据的知识管理可划分为知识生产、知识积累、知识交流和知识应用 4 个阶段，并形成知识反馈。其中，知识生产阶段包括知识发现环节和知识采集环节，知识积累阶段包括知识存储环节和知识检索环节，知识交流阶段包括知识分享环节和知识更新环节，知识应用阶段包括知识测评环节和知识服务环节，共 8 个环节。对于知识检索、知识搜寻、知识吸收、知识转移、知识忘记等环节，亦可归纳概括到上述 4 个阶段过程模型中。

传统知识管理过程是以知识搜寻为起点，以提升创新能力为目标的组织学习过程，而大数据时代的知识管理，其内涵与外延均有所扩展，由于数据特征的变化和技术手段的变革，呈现出差异性与复杂化。参考赵蓉英等[3]、孟小峰等[4]、许立波等[5]、叶英平等[15]的研究，结合大数据背景下知识管理技术工具的使用，本研究旨在大数据背景下，构建具有普遍适应性的知识管理模型（见图 4）。结合大数据环境分析具体的知识管理过程，从海量数据中提炼信息、发现知识、提供行为决策依据，最终实现商业利润与经济效益。同时，分析大数据特征为各个阶段的知识管理活动带来的挑战。值得注意的是，有些组织的知识管理过程可能呈现出跳跃式发展，跨过了其中的部分阶段；而有些组织的知识管理过程可能会出现反复式发展，在其中部分阶段循环进行，但大部分组织的知识管理过程，仍然遵循图 4 所示的知识管理模型的规律。为将知识管理模型与实际案例相联系，以如何提高百度知识搜索效率及精准度为例，对知识管理模型加以阐述。

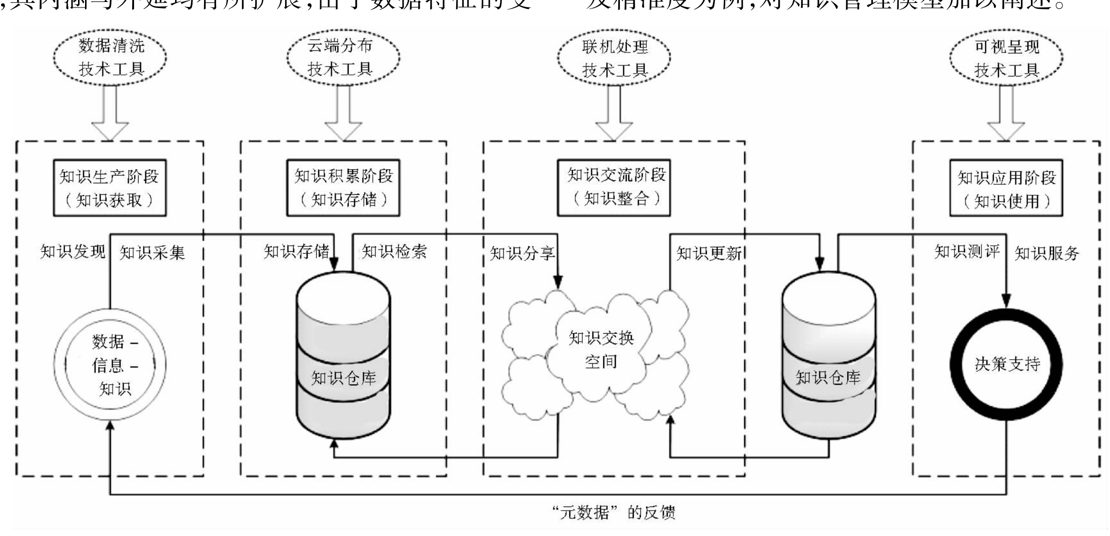


图 4　知识管理模型构建

4.1　知识生产阶段

传统知识管理的首要任务是知识搜寻，只有获取了一定量的知识，才能对这些知识进行后续的整合与吸收，而在海量碎片化信息的大数据时代，数据可以来源于互联网、物联网以及大量移动智能终端设备，知识获取的难度不在于信息搜寻，而在于如何从庞杂的数据中通过知识发现完成知识采集，如语义多样性即为百度搜索提出了这样的挑战。对于同样的知识信息，其对应的数据形式却未必一致，同时，在不同的语义情境下，相同的词句也可能表达不同的含义。为此，百度平台与其他搜索引擎相互合作，扩大了数据搜寻范围，同时，通过正则表达式的方式，将输入数据格式进行规范化处理，并将数据转化为信息，进而提炼、生产对应的知识，并将客户每次搜索点击的路径生成衍生数据。由于数据量大造成低密特征，为避免知识被淹没在大数据中，需要借助降噪手段使大数据缩水，从而在数据海洋中发现和采集更多知识。在知识生产阶段，通过知识发现和知识采集，可以结合大数据技术对市场进行扫描与监测，识别客户需求及潜在客户，获得更多客户知识资源。

4.2　知识积累阶段

知识生产阶段产出的知识是分散、零碎的，少量知识不足以支撑产品创新与技术进步，因而需要将这些知识积累在知识仓库中。经过清洗降噪预处理过程的知识在进入仓库后，可实现知识存储和知识检索功能。然而，先后进入仓库的知识往往会相互关联和连接，并且这种相关性或连接性会随着知识积累的增多而越来

越密切,导致知识之间的关系越来越复杂,为后续群体事件预测、关系变化建模等造成困扰。为此,百度搜索词条在知识积累的过程中对知识价值进行提纯,并建立知识间的关联性,运用知识图谱的方式,从海量数据库中为客户提供精准推送。区别于传统知识管理中的静态知识库,知识仓库的存储具有动态变化的特征,以此提升知识的实时索引能力。在知识积累阶段,通过不断积攒客户知识及市场知识,与组织原有知识储备形成互补性资源,能够有效弥补组织的知识短缺,使知识体系更加完善。

4.3 知识交流阶段

当面临知识交流需求时,可将知识仓库中的知识进行分享,但出于对知识源的产权保护以及对知识开发的多重可能,需要建立知识交换平台,将脱敏后的知识在独立的交换空间内进行分享,并根据所分享的新知识,补充、更新原有的知识体系。为提供给客户个性化的搜索服务,百度公司将使用同一 IP 地址及同一登录账户的多个终端设备的数据信息进行关联存储,并与设备在其他平台的浏览记录与交易信息进行共享存储和交叉分析。在知识交流阶段,通过知识分享能够实现组织的知识创新,对于共享知识,做到有则更新、无则创新,将新知识与原有知识相互融合,并与组织的实际经营情况相联系,切实推进组织产品、服务、经营模式等方面的知识创新。

4.4 知识应用阶段

在知识应用阶段,基于大数据情境的知识测评不同于传统商业调研,能够以知识分析和知识挖掘的方式,为组织提供全面、客观、详细的测评结果,辅助组织管理者做出正确的战略决策。此外,基于知识评价数据,可以对知识价值和知识关系做出测评,从而筛选出无用、重复、过时的知识,构建有序的知识网络;也可基于知识测评服务,为客户提供知识产品和决策咨询,即实现对外知识服务的目的。如百度公司根据索引的点击量对词条进行竞争性排序。在组织完成知识生产、知识积累、知识交流、知识应用4个知识管理过程后,对知识应用的评价数据会在知识发现与知识采集的环节重新被捕获,从而形成知识反馈,使知识在反复循环的过程中实现螺旋式上升的增值过程。在组织进行决策行为与经营活动的过程中,会产生关联数据与行为数据(这些关于数据的数据统称为"元数据"[4]),这些元数据会伴随着大数据洪流重新涌向组织,同时也面向所有互联网主体开放,伴随知识发现与知识采集,反复循环参与到知识管理的各个阶段。通过知识管理的反馈过程,能够实现有针对性的精准推送,以智能化的方式分析客户行为数据,挖掘客户需求规律,提供个性产品推荐及服务信息匹配,向客户传递所需产品及知识,从而增加客户对产品或服务的认知和需求,获得超额商业利润。可以说,从大数据到知识管理,既是逐步提升的螺旋上升结构,也是循环往复的转换过程。

5 大数据时代知识管理对策

大数据时代的知识管理是借助数据分析和数据挖掘的技术手段,将大量碎片化数据转化为有价值的知识,并为社会经济发展提供知识服务的过程。可见,原始数据、技术平台、应用价值是实现大数据时代知识管理的3个关键要素,此外,为保障知识管理的有序进行,信息安全问题也同样值得关注。为解决上述问题,需提出大数据时代知识管理的相关对策。

5.1 原始数据的碎片整合问题及对策

大数据时代的海量碎片数据将信息分割为非连续的、无逻辑的零碎单元,而这些多源异构的数据信息推动了当代知识传播在思维、表达、内容以及主体等方面的碎片化。任福兵[22]和张文德等[23]都曾指出,冗杂信息的存在使得在知识管理过程中,知识获取的时间维度和空间维度具有随机性与不确定性,即知识获取与传送在零碎的时间和多变的空间内完成。同时,数据的快速产生与更新也对知识管理的时效性和应激性提出了更高的要求。为应对大数据时代的知识管理,需要组织具有更高的变革思维,适时变化的沟通能力、分析能力和决策能力,有助于适应"碎片化"环境,更好地发挥知识管理的作用。大数据时代的知识传播形式向"言简意趣"的方向转变,这些用户生成内容对于相似观点的表达却广泛而杂碎,因而使得碎片化信息的获取更加复杂耗时,而知识的整合涉及多个类型、不同领域之间的交叉碎片信息。为此,通过知识仓库的建立,能够快速、完整地得到海量关联数据,将大数据下冗杂的碎片化信息进行整理,建立系统性的分析原则,构建所需知识框架、构造人机交互平台。通过系统化、结构化的知识库形成统一碎片化处理规则,增添访问功能的查询模块,设置用户需求权限,从而提高碎片数据的获取速度、提升信息整合的精准程度。

5.2 决策辅助的应用价值问题及对策

知识管理的目的在于知识服务,大数据背景下的知识管理不仅仅局限于对历史经验的梳理,更重要的是能够通过海量数据建立足够大的样本库,涵盖生产生活的各个领域。基于马尔可夫链的计算,能够推测

某一事件的发生概率,并推演未来发展的最可能方向,这种基于大数据分析的信息预测是客观科学的,并且随着机器学习的深度发展,其预测的准确率将不断提升。大数据背景下的知识管理正在逐渐摆脱信息不完全的束缚,因此可以应用于组织决策辅助领域,为客户提供决策支持服务。郭熙铜等[24]将数据挖掘引入了电子健康服务及医疗诊断决策辅助,胡小荣等[25]将数据分析应用于上市公司的风险识别与决策辅助,B. Gu等[26]将数据驱动应用于客户满意度评价,以替代传统的问卷调查和客户回访。知识管理能够辅助决策,相应地,决策辅助系统也能应用于组织知识管理行为的改进。

5.3 技术平台的硬件支撑问题及对策

伴随着互联网的发展,更多软件工具被应用于知识管理过程,如微软的 One Note 笔记本等知识管理软件以及 Service Hot 等知识管理系统。相比之下,学术界与产业界对知识管理的硬件支持设施关注度较低,知识管理的硬件系统基础设施建设也相对不足。知识管理平台是由专家系统、知识库等技术工具构成的综合系统,与支持知识收集、加工、存储、传递和利用的各种基础设施构成,能够快速分析知识结构、共享知识、对纷杂的知识内容分门别类处理。在构造知识管理平台的过程中,仅仅依靠先进的数据分析与知识管理软件是不够的,基础设施同样至关重要。存储设备如果损坏,会造成平台信息的丢失和泄露;处理器的故障则会造成系统运行的瘫痪;在联机处理过程中,运行速度较快的设备需要花费时间等待其他设备的延迟反映,造成了运算能力及能源消耗的浪费,诸如此类潜藏的经济损失不可估量。因此,为促进大数据背景下知识管理能力的提升,需要引进与软件相匹配、高效的基础硬件设施,并及时对硬件设备进行更新与维护工作,制定突发事件应急措施,当出现危机情形时,确保设备的基本运转以保障数据和信息的安全。

5.4 信息安全的隐私伦理问题及对策

大数据在对社会发展以及人类生活产生巨大改变的同时,也带来了不可忽视的数据隐私问题和信息安全风险。黄国彬等[27]认为隐私泄露的风险是由不同数据来源之间的组合、多种渠道信息的叠加,增加了对用户身份信息特征的推断概率;王世伟[28]也认为大数据分析技术如聚类、情感分析的广泛应用,正在逐渐获取用户的个人特征,窥视用户的私人生活,而这些用户隐私信息的泄露可能会带来误导、诈骗等社会问题,影响社会稳定。大数据背景下的知识内容是庞杂的,很多时候获取的信息并不能实现完整的利用,因此,在实际操作中,只获取学术研究与产业发展所需要的数据信息成为破解隐私保护问题的关键。在数据清洗与脱敏阶段,需要全面分析需求目标与预期结果,规划获取信息的规模、结构和层次等要素;在获取或者交换信息时,精准对接所需的数据结构,扩大有效信息占比。此外,仍然需要采取一定技术手段保护信息安全,如防火墙技术以及匿名保护技术的使用,能够有效监管与检测恶意信息获取行为,维护数据安全;同时,隐藏用户的身份信息以及行为特征,提高用户隐私保护的意识行为。

6 结语

大数据时代的到来意味着更加精准的知识管理与更加智能的知识服务,也为知识管理带来诸多机遇:①知识管理的系统化,由机器自动搜集的海量数据和传统人工进行的知识搜寻相比,虽然其价值密度有所降低,但因数据的信息量呈几何增长,所蕴含的价值总量仍然非常巨大,且价值体系更加丰富与完整;②知识管理的智能化,结合机器学习的智慧发展,能够将人类从知识管理的中低端解放出来,专注于高级知识的生产和应用;③知识管理的科学化,组织依靠大数据分析进行的运营决策更加客观和准确,通过数据挖掘还可以辅助组织预测未来发展方向,从而做出更加科学的战略规划。

大数据在提供种种机遇的同时,也逐渐暴露出一些可能的危害,为知识管理带来挑战:①数据安全问题,海量数据的开放会带来数据泄露和隐私安全的挑战,但信息过度窄化又会影响数据分析的准确性,如何在数据清洗和数据脱敏的过程中兼顾数据安全与数据价值,平衡数据公开与隐私保护的关系,是知识管理需要关注的问题;②大量异构数据的处理问题,知识管理需要应对数据量的多样化爆炸式增长,尤其是图像视频等格式的非结构化信息的大量涌现,对数据的存储与处理带来了挑战,如何以大数据思维重构组织业务流程,提高数据分析的准确性与及时性,是知识管理需要解决的问题;③硬件基础问题,日益复杂的硬件环境及其异构性导致在联机处理的过程中,大量时间浪费在等待性能较差的服务器做出计算,这种木桶效应为大数据的基础设施建设带来挑战,如何提高硬件集群的整体性能,是知识管理需要重视的问题。

随着数据量的不断增大,数据中心的存储规模也在不断扩张,高耗能将逐渐成为制约大数据发展的瓶

颈因素，如何在优化知识管理、提升知识服务的同时，降低大数据分析处理的能源消耗，是未来需要破解的资源难题。然而这方面尚未引起社会的足够重视，国内外相关学者往往更关注大数据的实践应用研究，因此，将大数据与其他技术融合发展，提高知识资源的生产与供给能力，同时，将大数据技术与小样本分析结合使用，也是大数据思维与大数据产业的重要融合手段。大数据并不是一个充斥着算法和机器的冰冷世界，通过强化知识管理和知识服务意识，期待知识型社会走进全民大数据时代。

参考文献：

[1] 刘智慧，张泉灵. 大数据技术研究综述[J]. 浙江大学学报（工学版），2014，48(6)：957－972.

[2] COOPER P. Data，information，knowledge and wisdom[J]. Anaesthesia & intensive care medicine，2014，15(1)：44－45.

[3] 赵蓉英，魏绪秋. 聚识成智：大数据环境下的知识管理框架模型[J]. 情报理论与实践，2017，40(9)：20－23.

[4] 孟小峰，慈祥. 大数据管理：概念、技术与挑战[J]. 计算机研究与发展，2013，50(1)：146－169.

[5] 许立波，潘旭伟，袁平，等. 知识智能涌现创新：概念、体系与路径[J]. 智能系统学报，2017，12(1)：47－54.

[6] 刘洁璇. 高校图书馆知识管理中的数据治理[J]. 情报科学，2018，36(1)：108－129.

[7] 刘捷. 大数据环境下基于知识管理的国有企业文档管理优化研究[J]. 科技情报开发与经济，2014，24(17)：121－123.

[8] 曾润喜，王琳，杜洪涛. 基于知识管理视角的大数据研究网络与结构研究[J]. 情报学报，2016，35(11)：1173－1184.

[9] 野中郁次郎，竹内宏高. 创造知识的企业——日美企业持续创新的动力[M]. 北京：水利水电出版社，2006.

[10] SIMONET A，FEDAK G，RIPEANU M. Active data：a programming model for managing big data life cycle across heterogeneous systems and infrastructures[J]. Future generation computer systems，2015，53(9)：25－42.

[11] 秦殿启. 整合与大数据理念下的个人知识组织[J]. 情报理论与实践，2014，37(2)：19－22.

[12] LIEW C S，WAH T Y，SHUJA J，et al. Mining personal data using smartphones and wearable devices：a survey[J]. Sensors，2015，15(2)：4430－4469.

[13] 黄天恩，孙宏斌，郭庆来，等. 基于电网运行仿真大数据的知识管理和超前安全预警[J]. 电网技术，2015，39(11)：3080－3087.

[14] 何军. 大数据对企业管理决策影响分析[J]. 科技进步与对策，2014，31(4)：65－68.

[15] 叶英平，卢艳秋，肖艳红. 基于网络嵌入的知识创新模型构建[J]. 图书情报工作，2017，61(7)：102－110.

[16] 彭宇，庞景月，刘大同，等. 大数据：内涵、技术体系与展望[J]. 电子测量与仪器学报，2015，29(4)：469－482.

[17] 程学旗，靳小龙，王元卓，等. 大数据系统和分析技术综述[J]. 软件学报，2014，25(9)：1889－1908.

[18] 金澈清，钱卫宁，周敏奇，等. 数据管理系统评测基准：从传统数据库到新兴大数据[J]. 计算机学报，2015，38(1)：18－34.

[19] 邱东. 大数据时代对统计学的挑战[J]. 统计研究，2014，31(1)：16－22.

[20] 任磊，杜一，马帅，等. 大数据可视分析综述[J]. 软件学报，2014，25(9)：1909－1936.

[21] 官思发，朝乐门. 大数据时代信息分析的关键问题、挑战与对策[J]. 图书情报工作，2015，59(3)：12－18.

[22] 任福兵. 碎片化与拼图化：网络传播的扩散与整合[J]. 情报资料工作，2014，35(3)：18－24.

[23] 张文德，程涵，刘田. 面向用户决策的高校信息碎片化整合模型[J]. 情报理论与实践，2018，41(3)：64－67.

[24] 郭熙铜，张晓飞，刘笑笑，等. 数据驱动的电子健康服务管理研究：挑战与展望[J]. 管理科学，2017，30(1)：3－14.

[25] 胡小荣，姚长青，高影繁. 基于风险短语自动抽取的上市公司风险识别方法及可视化研究[J]. 情报学报，2017，36(7)：663－668.

[26] GU B，YE Q. First step in social media measuring the influence of online management responses on customer satisfaction[J]. Production and operations management，2014，23(4)：570－582.

[27] 黄国彬，郑琳. 大数据信息安全风险框架及应对策略研究[J]. 图书馆学研究，2015，33(13)：24－29.

[28] 王世伟. 论大数据时代信息安全的新特点与新要求[J]. 图书情报工作，2016，60(6)：5－14.

作者贡献说明：

叶英平：撰写论文内容；

陈海涛：负责设计研究框架；

陈皓：负责资料收集、校对。

Knowledge Management Process，Technical Tools，Models and Strategies in the Period of the Big Data

Ye Yingping　Chen Haitao　Chen Hao

Management School of Jilin University，Changchun 130022

Abstract：［**Purpose/significance**］The period of big data has brought changes to traditional knowledge management. In order to adapt to the coming of the big data，it is necessary to re-understand the process of knowledge management，and use technology tools to build new knowledge management models to overcome the challenge of knowledge management.［**Method/process**］After the summarization of the formation process of data to knowledge，together with 4V fea-

tures of big data, and the analysis of big data technology tools needed for knowledge management, the knowledge management in the period of big data is divided into four stages: knowledge production, knowledge accumulation, knowledge exchange and knowledge application. With the help of technology tools, a new model of knowledge management is built. Solutions are proposed to the problems such as debris consolidation, application value, hardware support, and ethics on privacy. [**Result/conclusion**] Big data promote the development of the transformation of the knowledge management process from the traditional model. The model of knowledge management in the period of big data is closely related to the use of technology tools. It emphasizes the value of knowledge extracted from mass fragmentation data and more effectively assists organizational decision making. Therefore, hardware support, information security, and combining big data technology with small sample analysis will push knowledge management to a new level.

Keywords: period of big data technical tool knowledge management model